arXiv:2601.16236v1 [cs.SI] 21 Jan 2026

为网络中心性度量建立秩序

G. EXARCHAKOS、R. VAN DER HOFSTAD、O. NAGY 和 M. PANDEY

Abstract. 我们提出了一种定量方法，根据图中心性度量所诱导的顶点排序，比较任意一对图中心性度量。所提出的方法概念上简单、数学上优雅，并能以定量方式重新表述许多此前难以形式化的猜想。此外，该方法还提供了一种对网络科学家有用的近似方案。我们探讨了其中一些用途，并提出了若干本身也具有研究价值的新猜想。

1. Introduction

中心性度量在网络科学中的重要性，类似于导数或积分在微积分中的重要性。然而，尽管这两个概念对从业者而言都直观 易懂，但在一般情形下，其数学细节可能带来意想不到的复杂性，因此很难给出一个普遍 适用的数学定义。中心性度量的情形可以说不那么直接，因为有许多合理的中心性度量，能够捕捉网络中“处于中心”可能代表的不同含义。举例来说，该数据库[44] 目前列出了 400 多种中心性度量，每一种都有不同的应用场景作为依据。即使在近期，新的中心性度量仍在不断提出 [27]。这些不同的中心性度量并非只是同一理念的细微变体；通常，它们之间不存在易于处理的关系。关于中心性度量的全面介绍，请参阅 [59, 第 7]章和 [75, 14]。

本文提出了一种直接研究特定网络中各中心性度量彼此接近程度的方法。借此，我们通过提出一种定量方法，比较在同一图上定义的任意一对中心性度量，从而为中心性度量建立秩序。我们的方法基于这对中心性度量所诱导的顶点排序。直观而言，它反映了两种不同中心性度量所选出的排名靠前顶点集合之间的分歧。我们阐释了该方法，讨论其性质，并探讨了可能的应用场景。

背景。虽然中心性度量也有一些较为轻松的用途（例如，凯文·贝肯游戏 [32, 71] 与接近中心性有关；同样，厄尔德什″数 [39]用于衡量数学合作网络中的中心性），但它们也一直是各学科深入研究的主题。一个经典议题是现实世界网络的韧性、脆弱性及其他结构属性 [1, 3, 6, 9, 12, 20, 23, 30, 34, 40, 46]。此外，我们特别指出它们在交通网络 [21, 26, 41], 流行病学 [48, 65, 67, 76], 以及电力和电信网络的容错性 [25, 66, 69, 70]方面的应用。

中心性度量的数学性质也受到了广泛关注。PageRank 中心性度量由 [17, 62], 提出，据称符合 PageRank 幂律假设 [4, 5, 33, 35, 45, 50, 60, 61, 74], ，该假设指出，在大多数现实世界网络中，入度和 PageRank 遵循相同的幂律分布 [10, 24, 31, 54, 64]。令人惊讶的是，尽管这一假设往往成立，但最近已被证明不适用于有向优先连接模型 [5]。

中心性度量也受到统计物理学界的关注，因为相关问题往往可以借助统计物理的方法来研究。这一研究方向近期的一个例子是在 [8] 中使用腔体方法 。这一研究方向反复探讨的一个主题是

Key words and phrases. 中心性度量、比较、稀疏图、稠密图。1

引入新的近似方案，例如 [7], ，使从业者能够绕开某些中心性度量在计算复杂性方面的问题。我们推荐参阅 [68] ，以初步了解复杂性与近似问题。

比较中心性度量的研究常见于 [16, 36, 38, 57, 72], ，通常旨在探究在特定任务中，哪种中心性度量优于其他度量。例如， [57] 研究了根据不同中心性度量移除排名靠前的顶点对网络连通性的影响，从而比较各种中心性度量在识别现实世界网络中的连通性枢纽方面的能力。同样， [16]比较了各种中心性度量在从科学文章中提取关键词这一特定任务上的表现。除研究中心性度量之间的相关性外 [38], 几乎不存在不针对特定任务、对任意中心性度量进行一般性比较的研究。另一方面，也有人尝试根据中心性度量行为上的某些相似性对其进行聚类 [68], ，但这种聚类大多是启发式的，并未体现任何可量化的聚类标准。

关于各种中心性度量之间关系的研究并不多。在 [28], 中，给出了一种解析推导，说明了无向配置模型中的接近中心性与度中心性之间的关系。尽管在中心性度量中，我们主要关注其所产生的排序，而非精确公式，但这表明，顶点接近中心性的倒数与其度的对数呈线性关系。 [13]研究了 PageRank 对阻尼因子的依赖关系。

中心性度量之间可量化的关系也有助于计算。事实上，中心性度量的许多实际应用关注的是网络顶点的排序，而非度量的绝对值；通常，识别出中心性最高的顶点，比计算每个顶点的中心性更有用。

本文的主要创新。本文的主要创新在于，我们提出了一种定量比较不同中心性度量所产生排序的方法。为此，我们提出了所谓的中心性比较曲线（*CCC*），它描述了对于任意 $x \in (0, 1)$，网络中排名最高的 $x \cdot 100\%$ 顶点之间的交集，并且可以针对任意一对中心性度量进行计算。通过 CCC，可以立即看出各中心性度量是相似还是完全独立。由于该曲线是针对每个 $x \in (0, 1)$ 计算的，因此它展示了中心性度量在哪些方面一致、在哪些方面不同。当需要比较网络中中心性最高的顶点时，这尤其有用。

我们证明，CCC 具有若干理想性质。当两个中心性度量相等或互为单调变换时，CCC 恰好为 1；此外，CCC 不受单调变换的影响。这一性质很有用，因为我们主要关注的是中心性度量的排名。此外，当两个中心性度量相互独立时，CCC 接近于平方。最后，我们指出，在大规模图极限下，CCC 的表现良好：对于稀疏图，这适用于局部收敛；对于稠密图，则适用于图子极限。我们通过计算人工随机网络和现实世界网络中许多对中心性度量的 CCC，展示了 CCC 的用途，并考察哪些中心性度量对彼此接近、在排名靠前的顶点上接近，以及相互独立。就中心性度量之间可量化的关系而言，我们的实验表明，不同的中心性度量可能会产生相似的最中心顶点排序。这会带来计算方面的影响，因为此时可以用计算复杂性较低的度量，替代计算复杂性较高的度量。由此，我们的工作有望显著加快计算困难的中心性度量对最中心顶点的识别，不过这并非本文的主要关注点。

接下来，我们定义中心性比较曲线。

2. Results

2.1. 中心性比较曲线。设 $G = (V_G, E_G)$ 是一个图，其顶点集为 V_G，边集为 E_G。为简单起见，从现在起我们假设 $V_G = \{1, \dots, n\} = [n]$。

回顾一下，中心性度量是一个映射 $R\colon V_G \to \mathbb{R}_{\geq 0}$，其中 $R(v) > R(u)$ 应解释为：v 的中心性高于 u。我们可以将 R 视为一种对 V_G中的顶点进行排序的朴素方法。就我们的目的而言，我们需要细化这一排序，使其成为全序，并以一致的方式打破平局。为此，我们引入诱导顶点排序：

定义 1（诱导顶点排序）。固定一个图 $G = (V_G, E_G)$，令 R, S 为 G上的两个中心性度量。为每个顶点 $v \in V_G$ 赋予一个从标准均匀分布 Unif(0, 1) 中独立采样的数值 u_v。定义全序 $\prec_{(R,S)}$，使得对于任意 a, $b \in V_G$、$a \prec_{(R,S)} b$，当且仅当

（1）$R(a) < R(b)$；或（2）$R(a) = R(b)$，且 $S(a) < S(b)$；或（3）$R(a) = R(b)$, $S(a) = S(b)$ 且 $u_a < u_b$。

请注意，$\prec_{(R,S)}$ 与$\prec_{(S,R)}$不同。定义 1 中的全序使我们能够定义本文的主要贡献——中心性比较曲线：

定义 2（中心性比较曲线）。给定一个图 $G = (V_G, E_G)$，其有 n 个顶点。对于任意一个 V_G上的全序 $\prec$，将有序集（V_G, $\prec$）中的前 k 个元素记为 $\mathsf{Top}^G_{\prec}(k)$。取定义在 G上的任意两个中心性度量 R, S，并考虑定义 1 中诱导的排序$\prec_{(R,S)}$ 和$\prec_{(S,R)}$。对于 $A \subseteq V_G$，令 $|A|$ 表示 A中的顶点数。

大小为 $|V_G| = n$ 的图上的中心性比较曲线（CCC）是一个映射 $\mathsf{CCC}^G_{R,S}\colon (0, 1] \to [n]/n$，定义如下：

$$\mathsf{CCC}^G_{R,S}(x) = \frac{\left|\mathsf{Top}^G_{\prec_{(R,S)}}(\lceil xn \rceil) \cap \mathsf{Top}^G_{\prec_{(S,R)}}(\lceil xn \rceil)\right|}{n}. \tag{2.1}$$

换言之，$\mathsf{CCC}^G_{R,S}(k/n)$表示根据两个中心性度量 R 和 S得出的最具中心性的 k 个顶点的重叠，并按图的大小进行归一化。

在许多应用中，人们只关心两种中心性度量在排名最靠前的顶点集合上的一致程度。为此，我们引入一个由 CCC 导出的简单标量量，定义为

$$\mathsf{CCCo}^G_{R,S}(p) = \frac{\mathsf{CCC}^G_{R,S}(p)}{p}, \tag{2.2}$$

对于 p 较小的值，该量可以看作 $\mathsf{CCC}^G_{R在S}$ $x = 0$处的右导数近似值。一般而言，可以粗略预期，$n\mathsf{CCCo}^G_{R在S}(k/n)$中排名最靠前的 k 顶点里，约有 R 也会出现在 k 中排名最靠前的顶点里（对于 S）。例如，$\mathsf{CCCo}^G_{R在S}(0.05)$衡量两种中心性度量 R在 S 最具中心性的前 5% 顶点上的一致程度。该值越接近 1，说明二者越一致。

$\mathsf{CCC}^G_{R不S}$ 对图的边密度作任何假设；因此，它既适用于稀疏图和稠密图，也适用于有向图和无向图，从而成为一种普遍适用的比较方法。

2.2. CCC 的性质。本节将讨论中心性比较曲线（CCC）的若干性质。

Sy度量 y CCC。CCC 是 y对称的，因为 $\mathsf{CCC}^G{}_{R,S} = \mathsf{CCC}^G_{S,R}$.

CCC 的单调性。根据定义，$x \mapsto \mathsf{CCC}^G_{R,S}(x)$ 是非递减的。

对单调递增变换保持不变。 在文献中，中心性度量的定量比较通常通过计算中心性之间的相关性来进行。然而，相关性度量的是线性依赖关系，应用单调变换时，相关性可能会发生巨大变化。CCC 的优势在于，它对单调递增变换保持不变，因此更加稳健；也就是说，$\mathrm{CCC}^G_{R,\ S}$ 在同时用严格单调递增函数映射 R 和 S 时保持不变。因此，CCC 提供了一种定量方法，用于比较中心性度量之间的接近程度，以及网络中最具中心性的顶点；或者说，比较它们之间的差异。

当 $R = S$ 时，CCC 是恒等曲线。 如果 $R = S$，那么根据定义，不会发生重新排序，因此 $\mathrm{CCC}^G_{R,\ S}(x) = x$，这为完全一致提供了对角线参照。采用其他打破平局规则时，这一基本性质便不复存在，这也是我们制定该破同分规则的原因之一。此外，这也说明 $\mathrm{CCC}^G_{R,\ S}(x)$ 低于 x 的程度，衡量了 R 与 S 在排名最高的前 $\lceil nx \rceil$个顶点上有多不同 。

以恒等曲线为上界。根据定义，CCC 始终位于恒等线下方，即对于所有 $x \in [0,1]$，$\mathrm{CCC}^G_{R,\ S}(x) \leq x$ 。因此，如果 $\mathrm{CCC}^G_{R,\ S}(x)$ 接近 x，那么 R 与 S 大致相同。同样，如果在较小的x下，$\mathrm{CCC}^G_{R,\ S}(x)$ 接近 x，那么 R 与 S 将大致把相同的顶点排名为最具中心性的顶点。

差异最大的中心性度量的下界。显然，对于所有 $x \in [0,1]$，$\mathrm{CCC}^G_{R,\ S}(x) \geq 0$ ，而对于所有 $x \in [\frac{1}{2}, 1].$，$\mathrm{CCC}^G_{R,\ S}(x) \geq 2x - 1$。因此，对于满足$\mathrm{CCC}^G_{R,\ S}(x) \approx \max\{0, 2x-1\}$ 的中心性度量对 $R,\ S$，可认为它们差异最大。

使用 CCC 检验中心性度量的独立性。 此外，注意，如果 R 是一个随机的顶点排序，且独立于其他所有因素，那么当网络规模 $n \to \infty$ 时， $x \mapsto \mathrm{CCC}^G_{R,\ S}(x)$ 将收敛到函数 $x \mapsto x^2$ 。因此，当对所有 $x \in [0,1]$，$\mathrm{CCC}^G_{R,\ S}(x)$ 接近 x^2 时，我们可以认为这两个中心性度量完全无关。

比较总结。这些观察带来了自然的比较：如果 $\mathrm{CCC}^G_{R,\ S}(x)$ 接近恒等图，那么 $R,\ S$ 会产生非常相似的排名；而如果 $\mathrm{CCC}^G_{R,\ S}(x)$ 接近 x^2，这些排名则完全无关；当 $\mathrm{CCC}^G_{R,\ S}(x)$ 接近 $\max\{0, 2x-1\}$ 时，排名甚至会完全相反。

2.3. **将中心性比较曲线付诸实践**。我们在人工网络和现实世界的有向及无向网络上使用 CCC，以观察哪些中心性度量彼此相似，哪些并不相似。

对于人工网络，我们使用一个规模为 $[15,\ 58]$ $n = 100,\ 000$ 的有向配置模型，其入度和出度的幂律指数均为 3。我们使用配置模型的不同样本运行这项计算 100 次，并用红色误差条表示这 100 次运行均值的标准差。

我们考虑的一个重要现实世界网络是有向“hep-ph”引文网络 [52], ，它是网络科学中的基准数据集之一，最早在 [37]中使用。我们研究的其他现实世界网络包括由美国国家经济研究局维护、最早在 [51], 中研究的美国专利数据集，以及最早在 [53]中研究的斯坦福网络图。代码可在 https://github.com/nagyol/CCC 获取，完整的图集可在 https://doi.org/10.5281/zenodo.18214457 获取。有关这些计算的更多详情，请参阅补充材料。

We 将以 PageRank 为例，解释 CCC 的图示以及入度/出度 ee.

如何解读 CCC 图。 我们以 PageRank 与入度/出度为例，先说明如何解读 CCC 图。图 1 显示，在现实世界的引文网络中，PageRank 与入度相关，而与出度基本无关。第一个结论是基于这样一个事实：

图 1(a) 中蓝色 CCC 与虚线恒等曲线相当接近，而另一个结论成立，是因为图 1(b) 中的 CCC 几乎完美地贴合黄色 x^2 曲线。

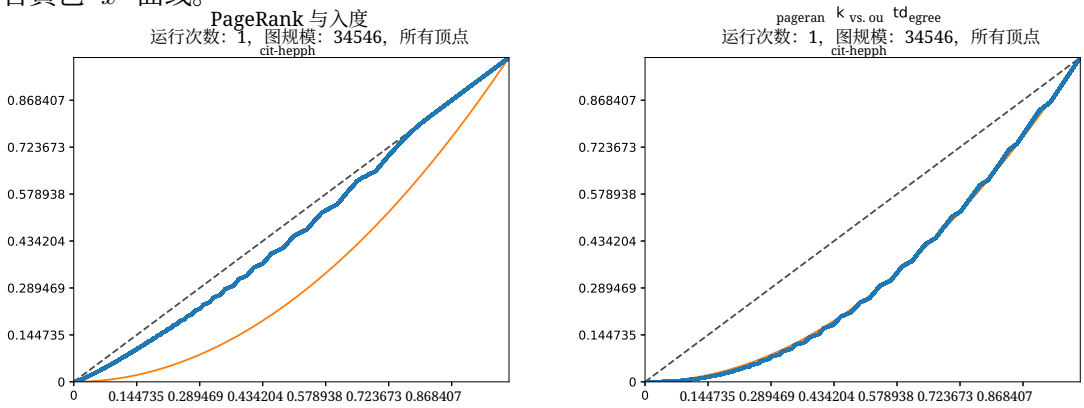

FIGURE 1. “hep-ph”引文网络中 PageRank 与入度、PageRank 与出度的 CCC [37]。

接近中心性与调和中心性，以及介数中心性与负载中心性相似。正如预期的那样，由于定义上密切相关，接近中心性与调和中心性通常较为接近（见图 2）。类似的关系也适用于介数中心性与负载中心性（见图 3）。

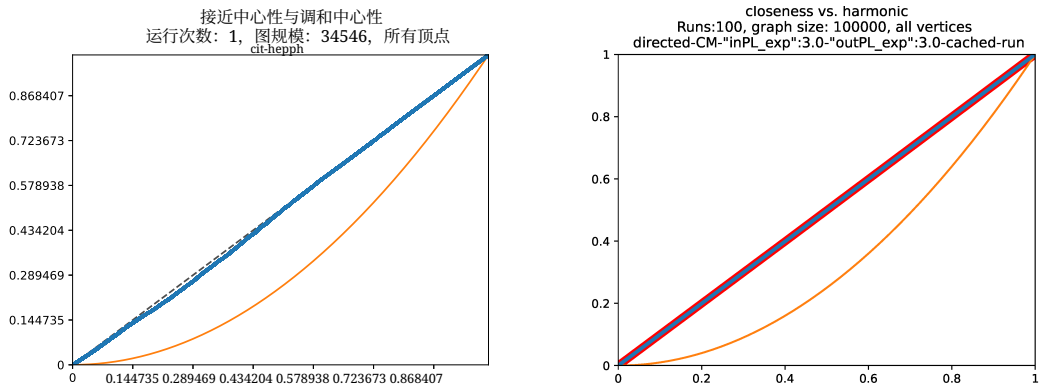

FIGURE 2. “hep-ph”引文网络 [37]和人工网络中，接近中心性与调和中心性的 CCC。

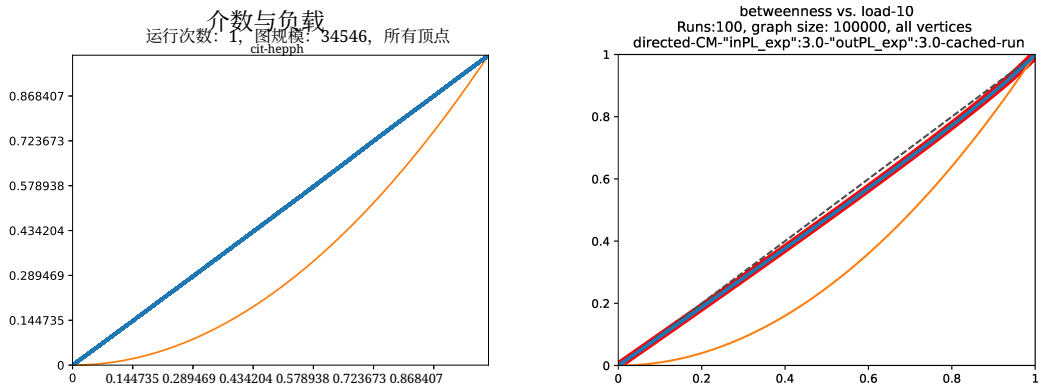

FIGURE 3. “hep-ph”引文网络 [37]和人工网络中，介数与负载中心性的 CCC。

PageRank 与入度的相关性较弱。入度中心性就是任意给定顶点的入度，而 PageRank 则是一种更复杂的随机过程，最初用于对 Google 搜索结果中的网页进行排序 [62]。大量研究表明，PageRank 分布与入度分布具有相同的幂律指数 [5, 10, 22, 24, 31, 50, 54, 64]。因此，毫不意外，这些中心性的 CCC 与一条直线相当接近（见图 4）。

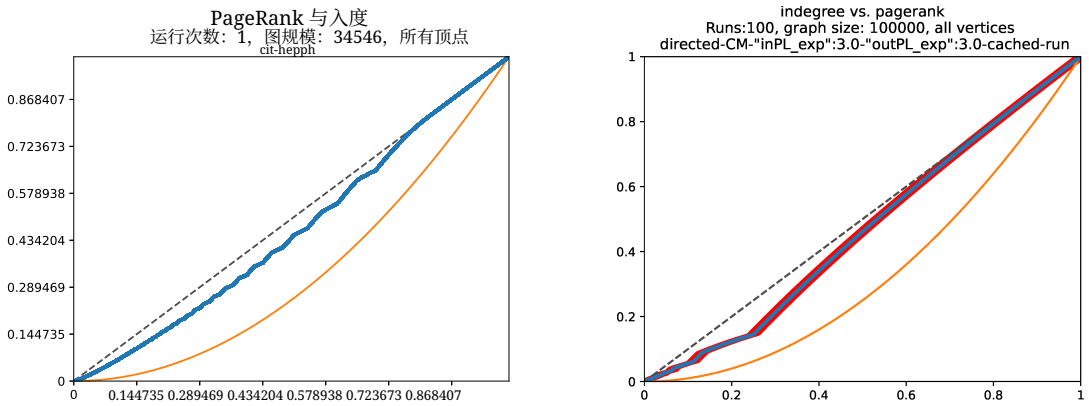

FIGURE 4. “hep-ph”引文网络和人工网络的 PageRank 与入度的 CCC[37]

PageRank 几乎不受阻尼因子的影响。PageRank 中心性有一个参数，称为 阻尼因子，通常取值为 0。85 [62]。不过，阻尼因子的具体取值并不重要，如图 5 所示。

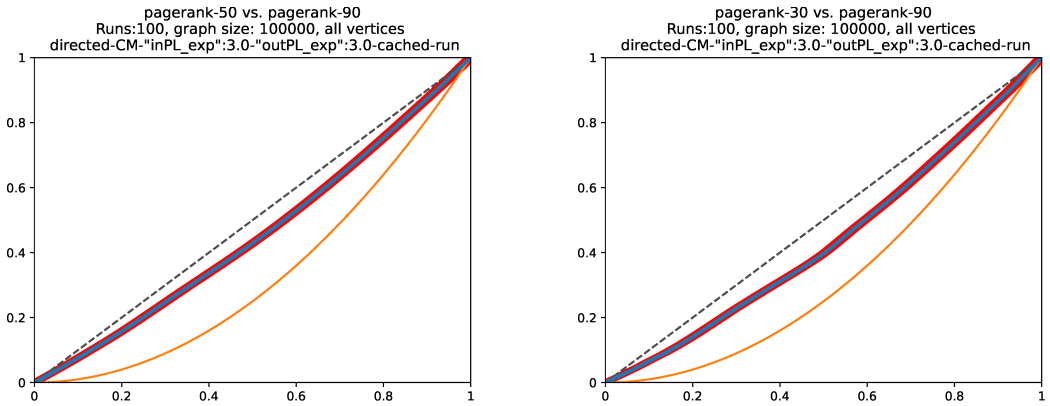

FIGURE 5. 人工网络中，阻尼因子分别为 0.5 与 0.9、0.3 与 0.9 时 PageRank 的 CCC。

Katz 与入度相似。有趣的是，在所考察的真实世界数据集中，Katz 中心性与入度几乎完全相同（见图 6 ）。

PageRank 与 Katz 有些不同。或许令人惊讶的是，尽管 PageRank 和 Katz 中心性的定义非常相似，但它们的表现却有所不同（见图 7）。

接近中心性与入度之间的关系较弱。在有向配置模型中，接近中心性与入度存在一定关联（见图 8）。

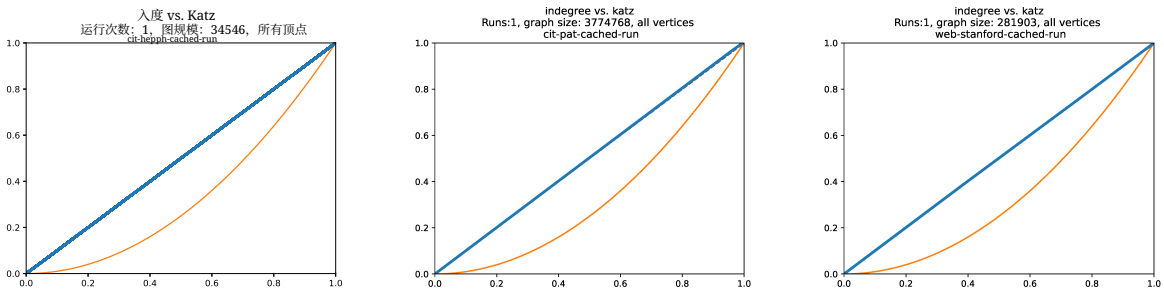

FIGURE 6. “hep-ph”引文网络中入度与 Katz 中心性的 CCC [37], “Patent”网络 [51], 以及斯坦福网页图数据集 [53]。

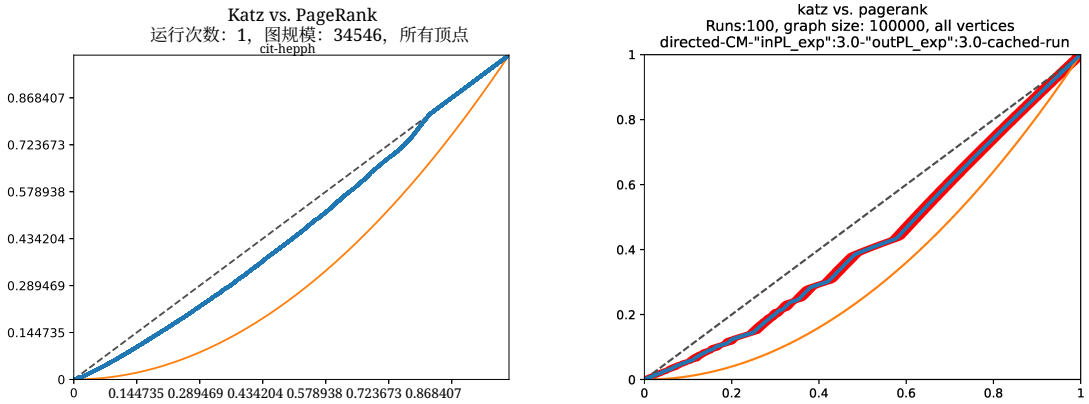

FIGURE 7. “hep-ph”引文网络中 PageRank 与 Katz 的 CCC [37]以及人工网络。

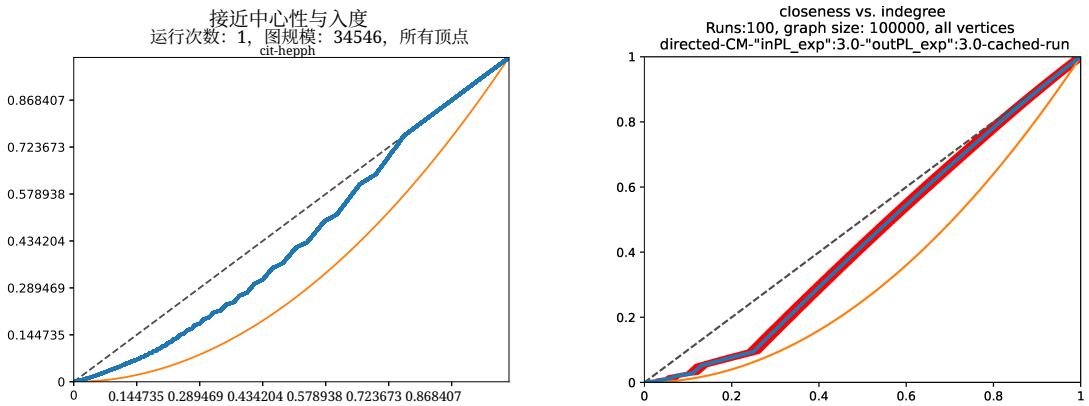

FIGURE 8. “hep-ph”引文网络中接近中心性与入度的 CCC[37]以及人工网络。

使用 CCC 提高计算效率。 CCC 可用于简化计算难度较高的中心性度量，例如介数中心性。图 6 展示了这一点：入度的计算显著比 Katz 中心性简单，但对于这些现实世界网络而言，两者几乎完全相同。另一个例子是，图 9 比较了介数中心性与基于长度至多为 6 和 10 的路径计算的介数中心性，分别记为

这里分别称为介数6和介数10，两者吻合度极高，尤其是在排名最靠前的顶点中。这使得我们能够在计算上难以处理的情形下，通过计算更简单的受限版本来近似介数中心性。例如，图10对比了介数6与介数10，显示两者吻合度极高；还对比了介数10与 PageRank。对于这个包含 3,774,768 个顶点和 16,518,948 条边的大型数据集，计算介数本身的开销相当大。

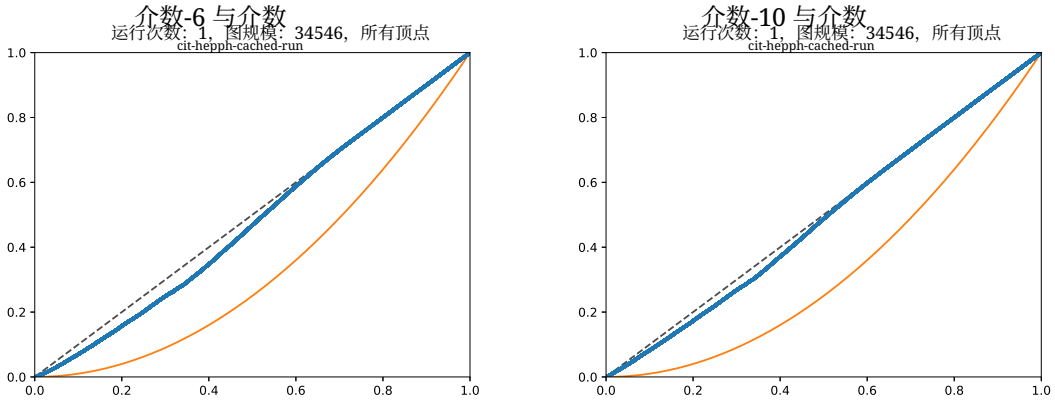

FIGURE 9. 介数与介数6、介数10的 CCC，分别针对“hep-ph”引文网络 [37]。

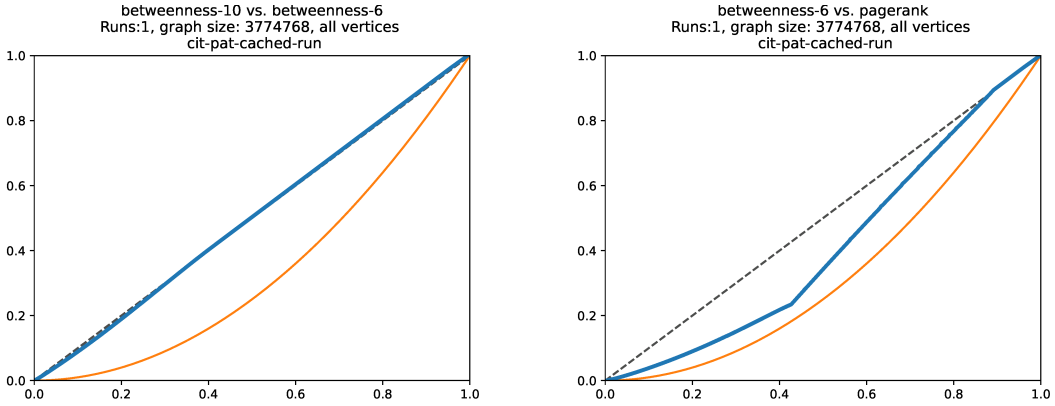

FIGURE 10. 介数6与介数10、以及介数6与 PageRank 的 CCC，分别针对“专利”网络 [51]。

2.4. 大图极限下中心性度量与 CCC 的收敛。 本节将讨论 CCC 在图序列的两种最重要收敛概念下的收敛：对于稀疏图序列，采用局部收敛 [2, 11] ；对于稠密图序列，采用图子收敛 [56, 77] 。

中心性度量的收敛以及局部中心性度量的 CCC。局部收敛 [2, 11], 和局部树状行为已成为网络科学中的核心技术。众所周知，PageRank 是一种局部量，因为顶点的 PageRank 可以通过有限半径邻域 [35]上的 PageRank 得到良好近似。显然，入度、出度和度中心性也满足这一点。根据其定义，我们还可以预期 Katz 中心性是一种局部度量。

对于成对的（联合）局部中心性度量，可以证明，CCC 会收敛到一条极限连续曲线 [29], ，该曲线可由图的局部极限确定。然而，局部度量并非全都相同（例如，参见图 7），尽管它们之间的关系可以通过局部极限进行分析。与此同时，对于稀疏图序列，许多全局中心性度量（如接近中心性、介数和负载中心性）无法预期会收敛，因为它们并非由局部邻域定义。不过，它们有时或许可以通过局部近似得到良好近似，如图 9–10 中介数的情况。

对于某些成对的中心性度量，其 CCC 在局部收敛图序列中会收敛。这一点可从我们的人工网络的 CCC 看出：这些网络是已知会局部收敛的随机图序列，而在不同模拟中，CCC 几乎没有变化。事实上，图中 CCC 的红线表示不同随机图模拟之间的变异程度，可以看出其变化相当小。这一特性十分重要，因为它表明 CCC 对相关图中的细微变化具有稳健 性，而这正是我们所期望的。例如，可以用基于长度仅为 k 的路径的全局 介数中心性度量替代局部 介数k 。如何有效运用这一方法的示例见图 9 和图 10。

中心性度量的收敛与稠密图序列的 CCC。 在稀疏图中，边数随网络规模线性增长。而在稠密图中，边数则随顶点数呈二次增长 。对于此类图序列，图子收敛是恰当的极限概念 [56, 77]。在许多情况下，对于收敛稠密图序列，接近中心性、介数中心性和负载中心性也会收敛，因此，此类中心性度量成对构成的 CCC 也会收敛。例如，对于具有独立边且严格 正的图子极限的稠密图，可以证明这一点。有关 Katz 中心性度量的图子极限及其对生产网络的影响示例，请参阅 [55] 。另一方面，对于（入度、出度）度中心性等中心性度量，每个顶点的度顺序在很大程度上取决于度波动，而图子极限无法反映这些波动。为边赋予适当的连续权重，并针对这类加权图序列计算中心性度量，或许可以缓解此类问题。

3. Discussion

本节将讨论我们的主要结果，并提出尚待解决的问题。

CCC 的性质。中心性比较曲线具有许多理想特性：它仅取决于中心性排名，因此不受单调变换的影响。当两种中心性度量一致时，它恰好是恒等曲线；当两种中心性度量完全独立时，它等于 $x \mapsto x^2$ 。最后，由于它是一条曲线，记录了两种中心性度量中排名最高的前 $x \cdot 100\%$ 顶点之间的一致程度，因此还可以据此判断这些度量在哪些排名群组中存在分歧。尤其是，即便两种中心性度量对大多数顶点的衡量差异很大，它们对排名最靠前的顶点仍几乎完全一致。对从业者而言，这往往最为重要。因此，它为比较成对的中心性度量提供了一种颇具吸引力的定量指标。

模拟得出的结论。CCC 表明，负载中心性与介数中心性密切相关，接近中心性与调和中心性也密切相关，这与它们各自的定义一致。此外，PageRank 与入度中心性相似，且基本不受阻尼因子选择的影响；而尽管 PageRank 与 Katz 中心性的定义相似，二者的接近程度却不如预期。

在无向情形中，定义相似的特征向量中心性和 PageRank 几乎表现得像彼此独立的中心性度量。同样，度和 Katz 中心性

在无向合作网络中似乎也各自发挥作用，成为彼此独立的中心性度量。相关图表请参见补充材料。

了解哪些中心性度量彼此相似，具有诸多应用。例如，对于相似的中心性度量，只需计算其中算法计算成本最低的一种即可。这也有助于检验由领域专家给出的直观排序是否与任何一种标准中心性度量高度吻合。最后，在动态网络中，可以通过绘制比较不同时间对的 CCC，观察中心性随时间的变化。

在文献中，通常通过计算中心性之间的相关性来比较中心性度量。然而，相关性度量的是线性 依赖关系；应用单调变换时，相关性可能会发生巨大变化。CCC 的优势在于，它在单调变换下保持不变 ，因此更加稳健。由此，CCC 提供了一种定量方法，用于比较中心性度量彼此的接近程度，以及网络中最具中心性的顶点；或者，也可以比较它们彼此独立的程度。

收敛图序列的 CCC 收敛性结论。CCC 对两种主要图收敛概念都具有良好的收敛性质，这一点颇具前景，也表明 CCC 很有用。同时，我们还需要进一步研究，才能充分发挥 CCC 收敛性的潜力。

CCC 的开放问题。 通过中心性比较曲线比较中心性度量，仍有许多尚待解决的问题；这里列出其中两个。

首先，可以对大量现实世界网络中的中心性度量之间的关系开展大规模调查。Broido 和 Clauset [18], 研究了度分布的尾部行为；此外， [73], 基于 KONECT 数据库 [49]调查了无标度网络究竟普遍存在还是相当罕见。上述论文得出了截然不同的结论，这恰恰说明此类分析颇具挑战性。对于中心性度量而言，需要量化不同度量在接近到何种程度时可视为接近；这既要考察全部顶点，也要考察排名最高的顶点，而这本身就并非易事。量化中心性度量之间的独立性意味着什么，也会很有帮助。接下来，还需要判断各种中心性度量对究竟是接近、相差甚远还是相互独立。这项工作应针对每个网络实例分别开展。最后，应对大量网络中的这些结论加以总结，从而全面了解哪些中心性度量对相似、哪些不同，以及哪些相互独立。

第二个开放问题涉及严格确立CCC 在稀疏随机图、稠密随机图以及介于两者之间的各种情形下的收敛性质。同样，这个问题包含多个方面。我们可以研究 CCC 在特定随机图模型中的收敛性，例如广受欢迎的网络模型，并考察这些模型上成对中心性度量的相似性。其次，更贴近我们在第 2.4 节讨论的结果类型的是，在大图极限下证明 CCC 对一般随机图模型的收敛性，这一点至关重要。有些实例表明，CCC 在稀疏图的局部极限和稠密图的图子极限这两种大图极限下表现良好；这是一个值得进一步研究的积极信号。

4. Methods: CCC and Graph limits

4.1. CCC 与全序。中心性得分通常只能形成偏序，因为顶点排名相同会导致并列排名。中心性比较曲线（*CCC*）基于排名，因此在采用全序时效果最佳。为了得到这样的全序，我们引入了一种结构化的破同分规则，同时保留 CCC 的可解释性。

假设 (R, S) 表示所考虑的两个中心性度量。在定义 CCC（回顾定义 1）时，首先根据主要度量 R 对顶点进行排名。随后按以下步骤打破平局：

(1) 首先使用次要度量 S 打破 R 中的平局。(2) 对于仍未分出先后的情况，则使用分配给每个顶点的相互独立的 Uniform(0,1) 随机变量打破平局。

所得的全序记为$\prec_{(R,S)}$，类似地，从 S开始排序会得到顺序$\prec_{(S,R)}$。

两条破同分规则背后的理由 是，步骤（1）确保使用 S 来解决$\prec_{(R,S)}$ 中的并列情况（对于$\prec_{(S,R)}$也对称地如此），从而最大化 两个排名中前k 个顶点集合的重叠。这保证 CCC 反映的是 R 和S 之间真正的差异，而非破同分造成的假象。接着，步骤（2）确保排序确实是全序，因为 Uniform(0，1)变量来自连续分布，因此以概率 1 打破任何剩余的并列。

全序的一个关键性质是，按照构造方式， $R = S$不会发生重新排序，因此 $\mathrm{CCC}^{G}_{R,\ S}(x) = x$，即代表完全一致的对角线，这是一项理想性质。在替代规则下，这一性质会丧失，如图 11(a)所示。相比之下，随机打破平局可能会夸大表观上的不一致，使 $\mathrm{CCC}^{G}_{R,\ S}(x)$更接近 x^2，即代表近似独立性的曲线。

最后，其他打破平局规则可以说不那么自然。事实上，当并列情况普遍存在时，纯粹随机打破平局会增加表观上的不一致；而与另一个度量无关的确定性规则，可能会人为减少重叠，却不能反映真实的结构差异。因此，分层打破平局规则（R， S，均匀分布）为通过 CCC 比较中心性排序提供了一个一致、可解释且对称的基础。

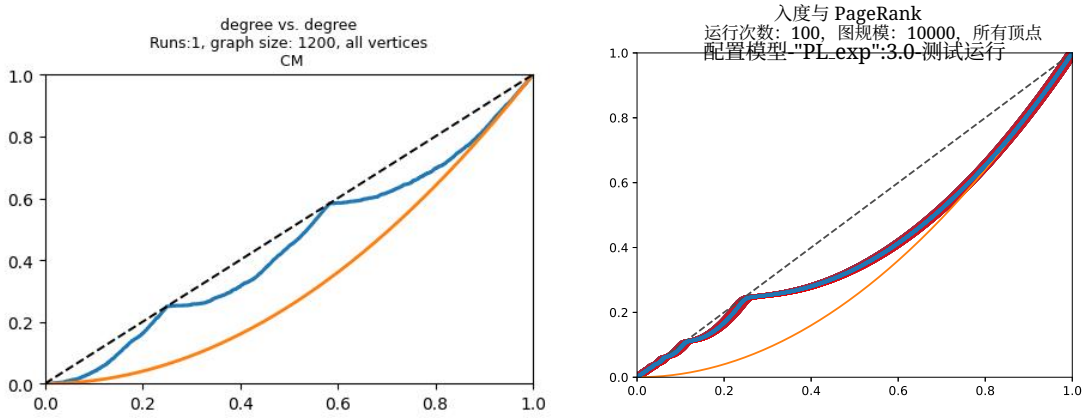

FIGURE 11. 替代打破平局规则对 CCC 的影响：对于一个包含 500 个度为 3、400 个度为 4 以及 300 个度为 5 的顶点的无向配置模型，采用随机打破平局规则时，度与度之间的 CCC；以及对于有向配置模型，采用替代打破平局规则时，PageRank 与入度之间的 CCC。请与图 4 进行比较。

4.2. **CCC 的一般性质**。 本节将说明第 2.2 节中讨论的 CCC 曲线若干性质背后的理由。首先，CCC 是对称的，因为其定义 (2.1) 对两个中心性度量是对称的。其次，CCC 曲线始终位于恒等线下方，这一点可直接由定义得出。此外，如果其中一个排序是随机的，那么 $\mathrm{CCC}^{G}_{R,\ S}(x)$ 表示按照两种排序，均出现在前 $x\cdot 100\%$ 中的顶点所占比例的期望值。当 n 较大时，该比例约为 x^2。当一个中心性度量对应于顶点的随机排序，而另一个则产生一种

确定性排序时，也会出现相同的情况；也就是说，该中心性度量不存在并列排名。这一观察说明，当 $\mathsf{CCC}^{G,\ s}_{R,\ S}(x)$ 接近 x^2时，将两个中心性度量称为 R 和 S 近似独立是合理的。

4.3. 稀疏图与局部收敛。 这里讨论稀疏图序列的局部收敛与 CCC 的收敛之间的关系。根图是一个二元组 $(G,\ v)$，其中 $v \in V_G$。如果图序列 $(G_n)_{n\geq 1}$ 满足：邻域形如某个特定根图的顶点所占比例依概率收敛到某个极限，则称该图序列局部弱收敛。这等价于以下表述：当 $o_n \in V_{G_n}$ 均匀随机选取时，随机根图 $(G_n,\ o_n)$依分布收敛到一个（随机的）极限根图，该极限根图可以是有限的，也可以是无限的。类似地，如果邻域形如某个特定根图的顶点所占比例依概率收敛到某个极限，则称图序列 $(G_n)_{n\geq 1}$ 依概率局部收敛。更多信息请参见 [43,第 2]章。

在本节及下一节中，我们将明确表示中心性度量对底层图的依赖关系，并用 $R^{(G)}(v)$ 表示图 G 中 $v \in V_G$ 的中心性。当 $R^{(G)}(v)$ 接近于在图 G 中顶点 v 的 r- 邻域 $B_r(v)$ 上计算的、顶点 v 处的中心性度量 $R^{(B_r(v))}$ 时，我们称中心性度量 $R^{(G)}\colon V_G \to [0,\infty)$ 局部。对于局部中心性度量，如果 $(G_n)_{n\geq 1}$ 在局部收敛，则 $R^{(G_n)}(o_n)$ 也依分布收敛于极限根图 (G,o) 的根中心性 $R^{(G)}(o)$。对于（入度、出度）中心性而言，这一点相对容易验证；但对于 PageRank [35], 也成立，并且在图的邻接矩阵的最大特征值满足某些条件时，对于 Katz 中心性 [63] 也成立。理解这一点的方法是证明：顶点 o_n 的邻域与作为附加于顶点的标记的中心性度量共同依分布收敛。当情况如此时，在温和条件下，顶点 o_n 的邻域与作为附加于顶点的双变量标记的局部中心性度量对也将共同收敛，这意味着 CCC 收敛。因此，对于成对共同局部的中心性度量，CCC 会收敛；这是一个相当好的性质。

显然，以上内容仅适用于成对的局部中心性度量，而许多全局（即非局部）中心性度量并不适用上述推理。重要的例子包括与图距离有关的中心性度量，如接近中心性、介数和负载中心性。同时，对于适当的图序列，这类中心性度量或许可以用局部量很好地近似（见图 9–10）。这带来了一个重大挑战，我们将留待进一步研究。

4.4. 稠密图与图子收敛。 接下来，我们讨论稠密图上的 $\mathsf{CCC}^{G,\ s}_{R,\ S}$ 的收敛性质。稠密图中的边数与网络规模的平方成正比。对于稠密图，自然的收敛概念是图子收敛 [56, 77]。在稠密图中，每个顶点 u 都有一个类型 x_u。我们假设 $(x_u)_{u\in V_G} \in [0,1]^n$ 满足其经验测度收敛于勒贝格测度。从顶点u到顶点 v存在一条有向边的概率为 $W(x_u,\ x_v)$，且各边相互独立。函数 W称为图子。我们考虑无向 情形，其中 W 是对称的；也考虑有向情形，其中 W 可能不对称，且 u 与 v 之间的边独立于 v 与 u之间的边。

令 $R = R^{(G)}$ 和 $S = S^{(G)}$ 为定义在图 G 上的两个中心性度量，其定义方式与上文基于图子 W 的定义相同。对于收敛的稠密图序列，若中心性度量 R 和 S 这一对的 CCC收敛 ，则意味着对于某个仅取决于图子 W的极限 C，有 $\mathsf{CCC}^{(G)}_{R\ S,}(x) \to \mathsf{C}(x)$。具体而言，这里我们关注（入度和出度）、接近中心性和介数中心性。如上文所述，在稀疏情形下，接近中心性和介数中心性并非局部量，因此这表明在稠密图中，收敛具有更广泛的适用性。

将范围限制在严格为正的图子 W 上颇有启发性，因为存在一个 $a>0$，使得对于所有 $x,\ y \in [0,1].$，都有 $W(x,\ y) \geq a$。还假设无向图 $(G_n)_{n\geq 1}$ 中的边是独立出现的。这样一来， G_n 中的距离表现得相当良好。事实上，以很高的概率， G_n 中的所有距离都为 1 或 2。因此，有 D_v 个顶点与

$v \in V_G$的距离为 1，而有 $n - D_v - 1$ 个顶点与 $v \in V_G$的距离为 2。因此，图中的距离与顶点的度密切相关。由此，v 的接近中心性也几乎由 D_v决定，介数和负载中心性等相关的、基于距离的中心性概念也是如此。由于度在图子收敛拓扑下是连续的，这应该也意味着 CCC 收敛；我们在模拟中清楚地看到了这一点，接下来将对此进行讨论。

我们模拟了规模不断增大的 CCC图；对于每种模拟规模，我们针对多个中心性度量对，在若干个图上计算 CCC，并从基于若干无向图子和有向图子的随机图中抽取 G 。为了模拟稠密图，我们取 n 个顶点，将它们的标签固定为 $\{1, \ldots, n\}$，并对每个顶点 v独立抽取一个标准均匀分布随机变量 x_v。然后，通常情况下，我们以概率 $W(x_i, x_j)$ 在两个顶点 i和 j之间独立添加一条边；对于某个 W，，该概率取决于具体设置，抽样过程见 [19]。

对于无向图子，我们使用乘积图子 $W(x, y) = cxy$，其中 $c < 1$。在图12中，我们可以看到所有中心性度量都完全相同；这一结论也适用于其他所有中心性度量对，相关图示可见 https://doi.org/10.5281/zenodo.18214457。此外，还可以在此处找到和图子 $W(x, y) = c(x + y)$ 对于 $c < \frac{1}{2}$ 的 CCC 图示，这些图示也都完全相同。

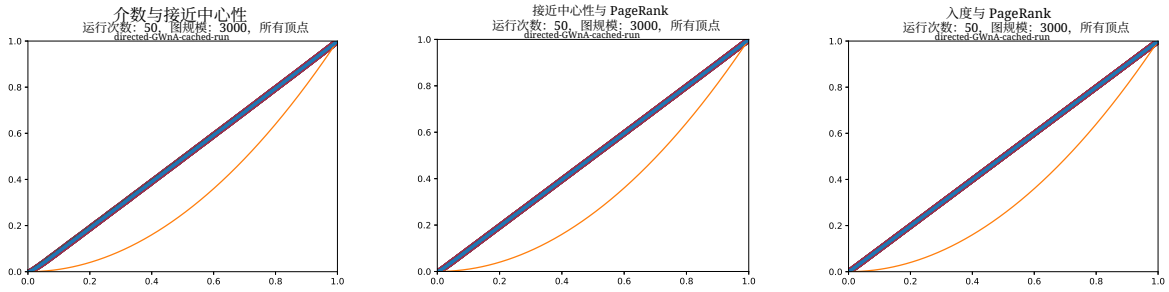

FIGURE 12. 无向乘积图子中接近中心性与介数、接近中心性与入度，以及入度与 PageRank 的 CCC。所有中心性均相同。

对于第一个有向图子，对于所有 $i < j$，从顶点 i 到j 存在一条有向边的概率为 $W_{01}(x_i, x_j) = cx_i x_j$，$c < 1$，从顶点 j 到 i $W_{10}(x_i, x_j) = c(1 - x_i)(1 - x_j)$，两者的入度和出度因此达到最大差异。结果见图

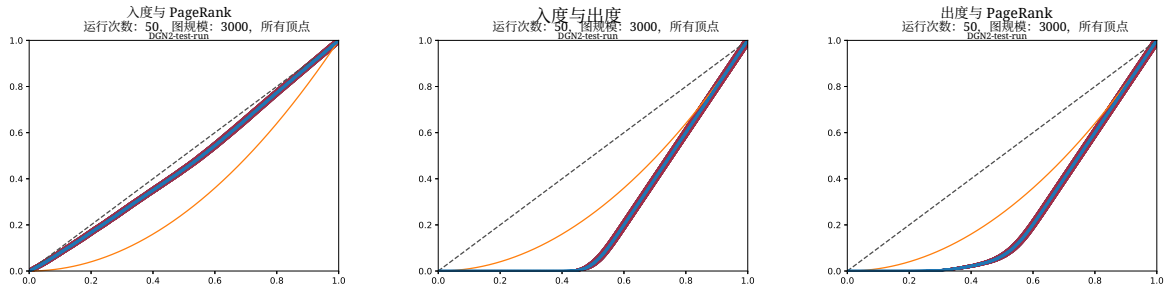

FIGURE 13. 有向乘积图子中入度与出度、入度与 PageRank，以及出度与 PageRank 的 CCC；该图子中的入度和出度差异最大。入度与 PageRank 几乎相同，而出度与入度和 PageRank 的差异接近最大。

对于 [19], 独立提出的有向图子，当 $x_j < p$时，从 i到 j 存在一条有向边的概率为 $W_{10}(x_i, x_j) = c_{\text{high}}$；否则为 c_{low}，其中常数 $c_{\text{high}}, c_{\text{low}}$和$p$ 分别取值为 $c_{\text{high}} = 0.9$、 $c_{\text{low}} = 0.05$和 $p = 0.15$。图14显示，PageRank与入度相似，而 PageRank 和入度与出度的差异都几乎达到最大。

我们得出结论，不同样本中 CCC 的变异程度很小，这表明大多数中心性度量对都确实具有收敛性。此外，对于无向图，我们发现 CCC 往往接近恒等曲线或最大差异曲线，这表明对于稠密图，许多中心性度量彼此相似，或者差异达到最大；这是一个重要的结论。

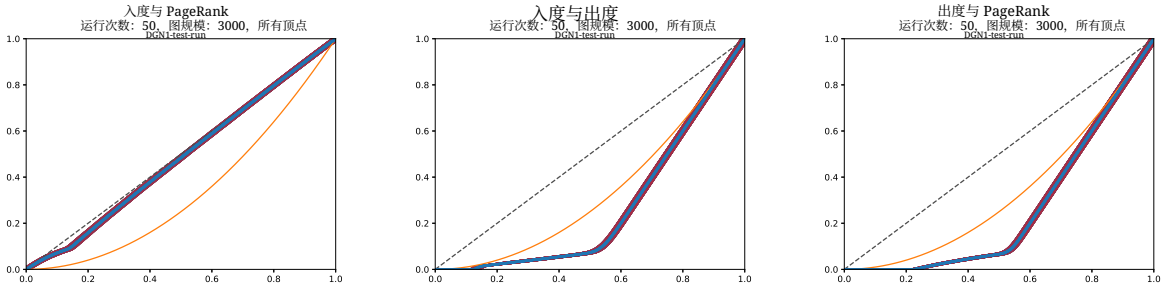

FIGURE 14. 入度与 PageRank、入度与出度，以及出度与 PageRank 的 CCC（针对 [19]）。

致谢。

本文的研究工作得到了荷兰科学研究组织（NWO）通过引力资助项目 NETWORKS-024.002.003 提供的支持。MP 的部分研究工作是在他此前任职于埃因霍温理工大学期间完成的。ON 的部分研究工作是在他此前任职于莱顿大学期间完成的。

REFERENCES

[1] R. Albert、H. Jeong 和 A.-L. Barab´ *Nature* 406 378 382, asi。复杂网络的错误与攻击容忍度。 ,: – 。2000 [2] D. Aldous 和 J. M. Steele。客观方法：概率组合优化与局部弱收敛。载于 *Probability on discrete structures*，《计算机科学讲义》第110卷，第1–72页。施普林格，柏林，2004年。[3] N. C. Asensio、E. M. Giner、N. S. de Groot 和 M. T. Burgas。宿主–病原体互作组中的中心性与感染期间的病原体适应度相关。 *Nature Communications*，8:14092，2017年。[4] K. Avrachenkov 和 D. Lebedev。无标度增长网络的PageRank。 *Internet Mathematics* 3 2 207 231, , ( ): – 。2006 [5] S. Banerjee 和 M. Olvera-Cravioto。有向优先连接网络上的PageRank渐近性质。 *Annals of Applied Probability* 32 4 3060, 2022 , ( ): 。[6] M. Bardoscia、P. Barucca、S. Battiston、F. Caccioli、G. Cimini、D. Garlaschelli、F. Saracco、T. Squartini 和 G. Caldarelli。金融网络的物理学。 *Nature Reviews Physics*，3:490–507，2021年。[7] S. Bartolucci、F. Caccioli、F. Caravelli 和 P. Vivo。利用聚合局部信息对网络中的有影响力节点进行排序。 *Phys. Rev. Research*，5:033123，2023年。[8] S. Bartolucci、F. Caravelli、F. Caccioli 和 P. Vivo。通过腔体方法研究无向随机网络上的中心性度量分布。 *arXiv preprint arXiv:2402.10566*，2024年。[9] S. Battiston、M. Puliga、R. Kaushik、P. Tasca 和 G. Caldarelli。债务排名：大到不能倒？金融网络、美联储与系统性风险。 *Scientific Reports*，2(1):1–6，2012年。[10] L. Becchetti 和 C. Castillo。只有在特定阻尼因子取值下，PageRank的分布才符合幂律。载于 *Proceedings of the 15th International Conference on World Wide Web*，第941–942页，纽约，2006年。ACM出版社。

[11] I. Benjamini 和 O. Schramm。有限平面图分布极限的递归性。 *Electron. J. Probab.*，6 23，13，2001 :第 .[12] S. Boccaletti、V. Latora、Y. Moreno、M. Chavez 和 D.-U. Hwang。复杂网络：结构与动力学。*Physics Reports*，424:175–308，2006年。[13] P. Boldi、M. Santini 和 S. Vigna。阻尼因子作为变量的PageRank。载于 *Proceedings of the 14th International Conference on World Wide Web* 05， 557 566， 2005 ，WWW ' 页面 – 纽约，NY，美国，。计算机协会。[14] P. Boldi 和 S. Vigna。中心性的公理。 *Internet Mathematics*，10(3-4):222–262，2014年。[15] B. Bollob´as。 *European J. Combin.* 1 4 311 316, 1980 as。关于带标号正则图数量渐近公式的一个概率证明。 , ( ): – 。[16] F. Boudin。基于图的关键词短语抽取中中心性度量的比较。载于 *Proceedings of the sixth international joint conference on natural language processing* 834 838, 2013 ，第 – 页。[17] S. Brin 和 L. Page。大型超文本网络搜索引擎的剖析。 *Computer Networks and ISDN Systems* 30 1 107 117, 1998 ，( ): – 。第七届国际万维网会议论文集。[18] A. Broido 和 A. Clauset。无标度网络十分罕见。 *Nature Commun.*，10(1):1017，2019年。[19] D. Cai、N. Ackerman 和 C. Freer。可交换有向图上的先验。 *Electron. J. Stat.*，10 3515，2016 2 3490( ): – 。[20] D. S. Callaway、M. E. J. Newman、S. H. Strogatz 和 D. J. Watts。网络鲁棒性与脆弱性：随机图上的渗流。 *Phys. Rev. Lett.*，85:5468–5471，2000年。[21] R. Carvalho、L. Buzna、F. Bono、E. Guti´errez、W. Just 和 D. Arrowsmith。泛欧天然气网络的鲁棒性。 *Phys. Rev. E*，80:016106，2009年。[22] N. Chen、N. Litvak 和 M. Olvera-Cravioto。有向配置网络上的广义PageRank。 *Random Structures & Algorithms* 51 2 237 274, 2017 ，( ): – 。[23] M. D. Domenico、A. Sol´e-Ribalta、S. G´omez 和 A. Arenas。随机故障下互连网络的可导航性。 *Proceedings of the National Academy of Sciences*，111:8351–8356，2014年。[24] D. Donato、L. Laura、S. Leonardi 和 S. Millozi。网页图的大规模特性。 *European Physical Journal* 38 239 243, 2004 ，: – 。[25] J. C. Doyle、D. L. Alderson、L. Li、S. Low、M. Roughan、S. Shalunov、R. Tanaka 和 W. Willinger。互联网"稳健却脆弱"的特性。 *Proceedings of the National Academy of Sciences*，102:14497–14502，2005年。[26] Y. Duan 和 F. Lu。不同粒度下城市道路网络的鲁棒性。 *Physica A: Statistical Mechanics and its Applications* 411 21 30, 2014 ，: – 。[27] M. Engsig、A. Tejedor、Y. Moreno、E. Foufoula-Georgiou 和 C. Kasmi。Domirank中心性通过节点支配性揭示复杂网络的结构脆弱性。 *Nature communications*，15(1):56，2024年。[28] T. S. Evans 和 B. Chen。关联接近中心性与度这两种网络中心性度量。 *Communications Physics*，5 1 2022 ( ):172，7月。[29] G. Exarchakos、R. van der Hofstad 和 M. Pandey。进行中的工作，2024+。[30] H. Farooq、Y. Chen、T. T. Georgiou、A. Tannenbaum 和 C. Lenglet。网络曲率是脑结构连接的标志。 *Nature Communications*，10:4937，2019年。[31] S. Fortunato、M. Bogu˜n´a、A. Flammini 和 F. Menczer。根据入度近似计算PageRank。载于 , , , , *Algorithms and Models for the Web Graph: Fourth International Workshop WAW 2006 Banff Canada November 30– December 1 2006 Revised Papers* 4936， 59 71， 2008 ，, , ,《计算机科学讲义》第 – 页，柏林，。施普林格。[32] B. Fowler。《凯文·贝肯六度关系》的真实历史，2019年8月。[33] N. Fraiman、T.-C. Lin 和 M. Olvera-Cravioto。有向随机图上的随机递归。 *Stochastic Processes and their Applications* 166 104055, 2023 ，: 。[34] J. Gao、B. Barzel 和 A.-L. Barab´asi。复杂网络中的普适韧性模式。 *Nature* 530 307 312, 2016 ，: – 。[35] A. Garavaglia、R. van der Hofstad 和 N. Litvak。PageRank的局部弱收敛。 *Ann. Appl. Probab.*，30 1 40 79，2020 ( ): – 。[36] V. Gaur、G. Soni、O. P. Yadav 和 A. Rathore。基于节点级联特性的网络中心性度量比较。载于 *Proceedings of 5th International Conference on Recent Trends in Computing: ICRTC 2020* 181 187 2021 ，第 – 页。施普林格，。[37] J. Gehrke、P. Ginsparg 和 J. M. Kleinberg。2003年KDD杯概述。 *SIGKDD Explorations*，5(2):149–151，2003年。[38] N. Ghazzali 和 A. Ouellet。社交网络中心性度量比较研究。载于 *Information Systems for Crisis Response and Management in Mediterranean Countries: 4th International Conference ISCRAM-med 2017 Xanthi Greece October 18-20 2017 Proceedings 4*，, , , , , 3 16 2017 ，第 – 页。施普林格，。[39] J. Grossman。厄尔多什数项目，2022年。[40] D. Guilbeault 和 D. Centola。用于识别和预测复杂传染传播的拓扑度量。 *Nature Communications*，12:4430，2021年。

[41] R. Guimer`a、S. Mossa、A. Turtschi 和 L. A. N. Amaral。《全球航空运输网络：异常中心性、社区结构与城市的全球角色》。 *Proceedings of the National Academy of Sciences*，102 7794 7799，2005 年：–。[42] A. Hagberg、P. Swart 和 D. S Chult。《使用 NetworkX 探索网络结构、动力学和功能》。技术报告，洛斯阿拉莫斯国家实验室（LANL），美国新墨西哥州洛斯阿拉莫斯，2008 年。[43] R. v. d. Hofstad。 *Random graphs and complex networks. Volume 2*，第 **54** *Cambridge Series in Statistical* 卷。剑桥 *and Probabilistic Mathematics* 2024 大学出版社，。[44] M. Jalili、A. Salehzadeh-Yazdi、Y. Asgari、S. S. Arab、M. Yaghmaie、A. Ghavamzadeh 和 K. Alimoghaddam。CentiServer：用于中心性分析的综合资源、网络应用程序和 R 软件包。 *PLOS ONE* 10 11 1 8, 11 2015 ，( )：–。

[45] P. R. Jelenkovi ´ *Advances in Applied* c 和 M. *Probability* 42 4 1057 1093, 2010 Olvera-Cravioto。树上的信息排名与幂律。，( )：–。[46] H. Jeong、S. P. Mason、A.-L. Barab´asi 和 Z. N. Oltvai。蛋白质网络中的致死性与中心性。 *Nature*，411 41：–42，2001 年。[47] L. Katz。基于社会计量分析得出的一种新地位指数。 *Psychometrika*，18(1)：39–43，1953 年。[48] M. Kitsak、L. K. Gallos、S. Havlin、F. Liljeros、L. Muchnik、H. E. Stanley 和 H. A. Makse。识别复杂网络中有影响力的传播者。 *Nature Physics*，6：888–893，2010 年。[49] J. Kunegis。KONECT：科布伦茨网络集合。载于 *Proceedings of the 22nd International Conference on World Wide Web* 1343 1350, 2013 ，页码 –。

[50] J. Lee 和 M. Olvera-Cravioto。非齐次随机有向图上的 PageRank。 *Stochastic Process. Appl.*，130 4 2312 2348, 2020 年 ( )：–。[51] J. Leskovec、J. Kleinberg 和 C. Faloutsos。图演化：致密化与直径收缩。 *ACM transactions on Knowledge Discovery from Data (TKDD)* 1 1 2 2007 ，( )：–

es，。[52] J. Leskovec 和 A. Krevl。SNAP数据集：斯坦福大型网络数据集集合。http://snap.stanford. edu/data，2014 年 6 月。[53] J. Leskovec、K. Lang、A. Dasgupta 和 M. Mahoney。大型网络中的社区结构：自然簇大小及大型明确簇的缺失。 *Internet Math.*，6(1)：29–123，2009 年。[54] N. Litvak、W. R. Scheinhardt 和 Y. Volkovich。入度和 PageRank：为何它们遵循相似的幂律？ *Internet mathematics*，4(2-3)：175–198，2007 年。[55] Z. Liu。连续生产网络中的Graphon均衡。可见 https://papers.ssrn.com/sol3/ papers.cfm?abstract_id=5788282，2025 年。[56] L. Lov ´ *Large networks and graph limits* **60** *American Mathematical Society Colloquium Publica-* asz。，第 *tions* 2012 卷。美国数学学会，罗得岛州普罗维登斯，。[57] D. C. Mocanu 、 G. Exarchakos 和 A. Liotta。对复杂网络中隐藏关系的去中心化动态理解。 *Scientific Reports*，8(1)：1571，2018 年 1 月。[58] M. Molloy 和 B. Reed。具有给定度序列的随机图中巨型连通分量的大小。*Combin. Probab. Comput.*，7(3)：295–305，1998 年。[59] M. Newman。 *Networks*。牛津大学出版社，牛津，2018 年。第二版。[60] M. Olvera-Cravioto。树上线性递推解的尾部行为。 *Stochastic Processes and their Applications* 122 4 1777 1807, 2012 ，( )：–。

[61] M. Olvera-Cravioto。度相关性下 PageRank 的行为。 *Annals of Applied Probability*，31(3)：1403–1442，2021 年。[62] L. Page、S. Brin、R. Motwani 和 T. Winograd。PageRank引文排名：为网络带来秩序。技术报告 1999-66，斯坦福信息实验室，1999 年。[63] M. Pandey *Centrality measures and connectivity properties in large networks: Who is the most influential in a network?* 2025 博士论文，埃因霍温理工大学，。[64] G. Pandurangan、P. Raghavan 和 E. Upfal。利用 PageRank 描述网络结构。载于 *Computing and Combinatorics: 8th Annual International Conference COCOON 2002 Singapore August 15–17* ，，， *2002 Proceedings* 330 339, 2002 ，页码 –。

[65] R. Pung、J. A. Firth、L. G. Spurgin、A. Chang、J. Kong、J. Wong、O. J. Jin、D. Selvaraj 等。利用高分辨率接触网络评估大规模多日活动中的 SARS-CoV- 2 *Nature* 传播与控 *Communications* 13 1 1956, 2022 制。，( )：。[66] S. M. Rinaldi、J. P. Peerenboom 和 T. K. Kelly。识别、理解和分析关键基础设施相互依赖关系。 *IEEE Control Systems Magazine*，21：11–25，2001 年。[67] M. Salath´e、M. Kazandjieva、J. W. Lee、P. Levis、M. W. Feldman 和 J. H. Jones。用于传染病传播研究的高分辨率人类接触网络。 *Proceedings of the National Academy of Sciences*，107：22020–22025，2010 年。[68] A. Saxena 和 S. Iyengar。复杂网络中的中心性度量：综述。 *arXiv preprint arXiv:2011.07190*，2020 年。

[69] B. Sch¨afer、D. Witthaut、M. Timme 和 V. Latora。电网中动态诱发的级联故障。*Nature Communications*，9：1975，2018 年。[70] R. V. Sol´e、M. Rosas-Casals、B. Corominas-Murtra 和 S. Valverde。蓄意攻击下欧洲电网的稳健性。 *Phys. Rev. E*，77：026102，2008 年。[71] I. Teotonio。Google 在搜索引擎中加入凯文·贝肯六度关系。 *Toronto Star*，2012 年 9 月。[72] K. Vignery 和 W. Laurier。中心性度量的方法论与理论分类：学生网络中哪些中心性指标最佳？ *PLOS ONE*，15(12)：e0244377，2020 年。[73] I. Voitalov、P. van der Hoorn、R. van der Hofstad 和 D. Krioukov。精心构建的无标度网络。

Phys. Rev. , ( )： 。

Res. **1** 3 033034, 2019

[74] Y. Volkovich 和 N. Litvak。个性化网页搜索的渐近分析。 *Adv. in Appl. Probab.*，42(2)：577–604，2010 年。[75] Z. Wan、Y. Mahajan、B. W. Kang、T. J. Moore 和 J. H. Cho。中心性指标及其网络韧性分析综述。 *IEEE Access*，9：104773–104819，2021 年。[76] Z. Wang、C. T. Bauch、S. Bhattacharyya、A. d'Onofrio、P. Manfredi、M. Perc、N. Perra、M. Salath´e 和 D. Zhao。疫苗接种的统计物理学。 *Physics Reports*，664：1–113，2016 年。[77] Y. Zhao。 *Graph theory and additive combinatorics—exploring structure and randomness*。剑桥大学出版社，剑桥，2023年。

APPENDIX A. SUPPLEMENTARY MATERIAL

A.1. 常用中心性度量的定义。 网络中的中心性度量是赋予顶点的数值，用以表示其重要性。现在我们正式定义中心性度量。设 $G=(V_G,E_G)$ 是一个图/有向图。中心性度量 $R\colon V_G\to\mathbb{R}_{\geq 0}$ 是一个从顶点集映射到正实数的函数，其在某个顶点处的取值表示该顶点的重要性。

文献中已有许多中心性度量。本文考察入度、出度、总度、PageRank、介数中心性、接近中心性、负载中心性、调和中心性、Katz 中心性和特征向量中心性；这些度量既适用于有向图，也适用于无向图。

PageRank。 设 $R\colon V_G\to\mathbb{R}_{\geq 0}$ 为 PageRank 函数，它满足以下递归关系

$$R(i)=c\sum_{j\in V_G}\frac{e_{j,i}}{d_j^+}R(j)+(1-c), \tag{A.1}$$

其中，$e_{j,i}$ 是 G 中从 j 到 i 的边数，d_j^+ 是 j 的出度，而 $c\in(0,1)$ 是阻尼因子。

Katz 中心性。 Katz 中心性 $K_\alpha:\ V_G\to\mathbb{R}_{\geq 0}$，由 Leo Katz [47]，提出，是一种定义如下的中心性度量：

$$K_\alpha(i)=\sum_{k\geq 1}\sum_{j\in V_G}\alpha^k A^k_{i,j}, \tag{A.2}$$

其中 A 是 G 的邻接矩阵，$\alpha\in(0,1)$ 是衰减因子 r.

请注意，为使 (A.2) 中的求和收敛，衰减因子 α 的取值必须小于 A 的最大特征值的倒数（该特征值由于 A 的非负性而非负）。

特征向量中心性。 令 $x\colon\ V_G\to\mathbb{R}_{\geq 0}$ 为特征向量中心性向量，则其满足

$$x^tA=\lambda x^t, \tag{A.3}$$

其中 A 是 G 的邻接矩阵，λ 是其对应的最大模特征值。该特征向量在乘以一个因子后是唯一的。为定义唯一的得分，必须对特征向量 x 进行归一化。

介数中心性。 设 $b\colon V_G\to\mathbb{R}_{\geq 0}$ 为介数中心性函数，则

$$b(v)=\sum_{i\in V_G}\sum_{j\in V_G}\frac{\sigma_{i,j}(v)}{\sigma_{i,j}}, \tag{A.4}$$

其中 $\sigma_{i,j}(v)$ 计数从 i 到 j 且包含 v 的最短路径数量，而 $\sigma_{i,j}$ 是从 i 到 j 的最短路径数量。对于介数k，上述讨论仅限于满足 i 和 j 的顶点对 $i,\ j\in V_G$，使得它们之间的图距离至多为 k。

负载中心性。 设 $l\colon V_G\to\mathbb{R}_{\geq 0}$ 为负载中心性函数，则

$$l(v)=\frac{\sum_{i\in V_G}\sum_{j\in V_G}\sigma_{i,j}(v)}{\sum_{i\in V_G}\sum_{j\in V_G}\sigma_{i,j}}, \tag{A.5}$$

其中，$\sigma_{i,j}(v)$表示包含 v 的从 i到 j 的最短路径数量，而$\sigma_{i,j}$ 表示从 i 到 j的最短路径数量。对于负载k，我们将上述范围限制为满足 i 且 j 之间的图距离至多为 k的顶点对 $i,\ j\in V_G$。

Cl接近中心性。令 c: $V_G \to \mathbb{R}_{\geq}0$ 为接近中心性函数，

$$c(v) = \frac{|V_G| - 1}{\sum_{u \in V_G} d_G(u, v)}, \tag{A.6}$$

其中，$d_G(u,\ v)$ 是 u 与 v 在 G中的最短路径长度。

Har单中心性。设 h: $V_G \to \mathbb{R}_{\geq 0}$ 为调和中心性函数，则

$$h(v) = \sum_{u \neq v} \frac{1}{d_G(u, v)}, \tag{A.7}$$

其中，$d_G(u,\ v)$ 是 u 与 v 在 G中的最短路径长度。

A.2. **所用人工网络的详细信息**。用于生成正文图表的源代码可通过 https://github.com/nagyol/CCC 在 GitHub 上获取。该程序广泛使用了 NetworkX 和 igraph 库 [42]。

模拟有向配置模型。对于度序列服从指数为 α的幂律分布的（有向或无向）配置模型，各个度数的选取方式为

$$\deg(v) = \lfloor d_{\min} + Y \rfloor,$$

其中，度的下界 $d_{\min}$ 取为 1，而 Y 是一个服从 Pareto($\alpha - 1$) 分布的随机变量，即其累积分布函数为

$$F_Y(x) = \begin{cases} 1 - \frac{1}{x^\alpha} & x \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

对于有向配置模型，首先通过上述流程对入度序列进行抽样，然后重复抽样出度序列，直到其总和与入度序列的总和相同。

模拟有向图子。对于基于图元的模拟，重点在于展示随着单个图的规模趋于无穷，CCC 的猜想极限。单个图按照以下流程进行抽样：

(1) 开始时创建一个包含 n 个顶点的空图。(2) 对于每个顶点 $v \in V_G$，分配一个权重 $x_v \sim$ Unif($[0, 1]$)。(3) 根据独立 Bernoulli(cx_ux_v) 试验的结果，在边 $u, v \in V_G$ 之间创建一条边。

这种随机图也称为 $G(W,\ n)$ 模型（例如参见 [56, 77]）。

A.3. **无向图的 CCC**。本节展示的图与论文正文中的图相同，但此处针对的是无向现实世界网络和人工网络。在这类网络中，中心性度量同样被广泛用于识别网络中的重要顶点。现实世界网络由 [51], 中研究的合作网络构成，而人工网络则对应于幂律指数为 3 的配置模型。

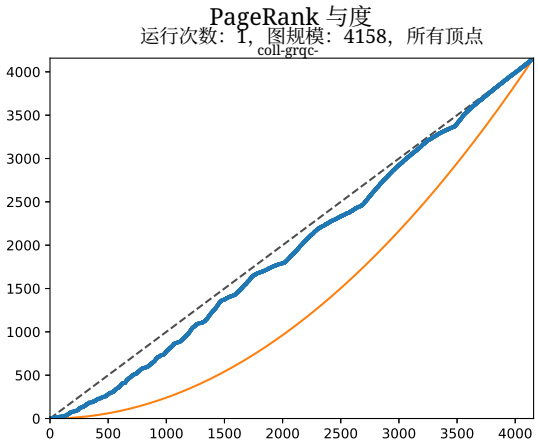

FIGURE 15. 合作网络 [51]中 PageRank 与度的 CCC。

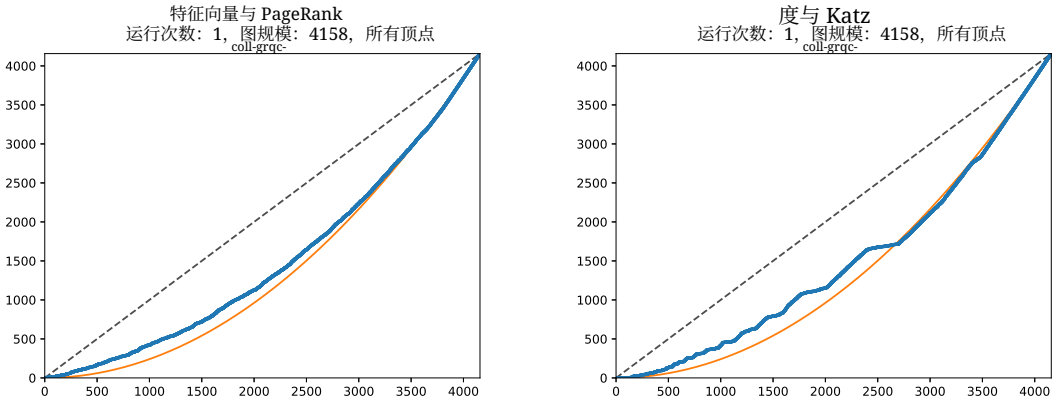

FIGURE 16. 合作网络 [51]中，特征向量与 PageRank，以及度与 Katz 中心性的 CCC。

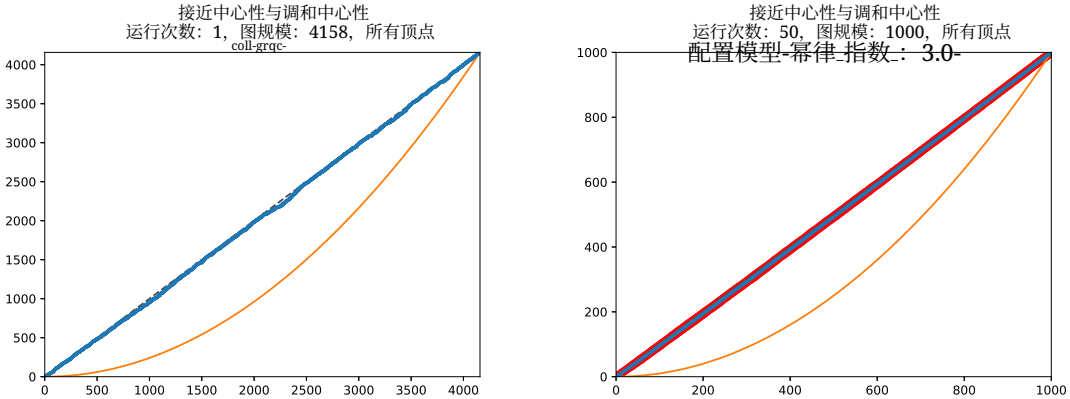

FIGURE 17. 合作网络 [51]和人工网络的接近中心性与调和中心性的 CCC。

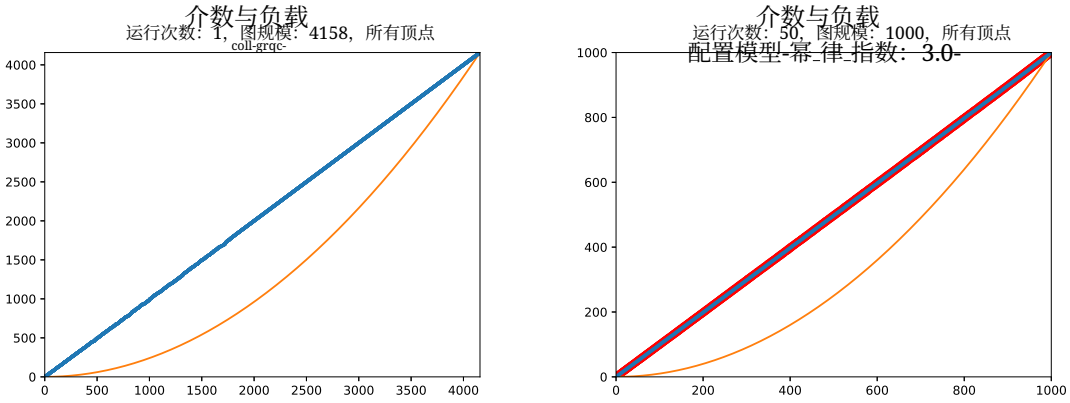

FIGURE 18. 合作网络中介数与负载中心性的 CCC [51]以及人工网络。

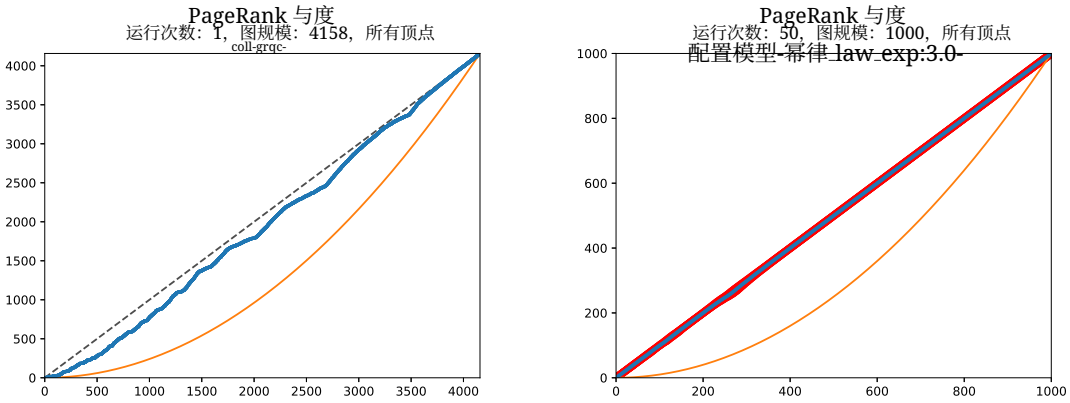

FIGURE 19. 合作网络中 PageRank 与度的 CCC [51]以及人工网络。

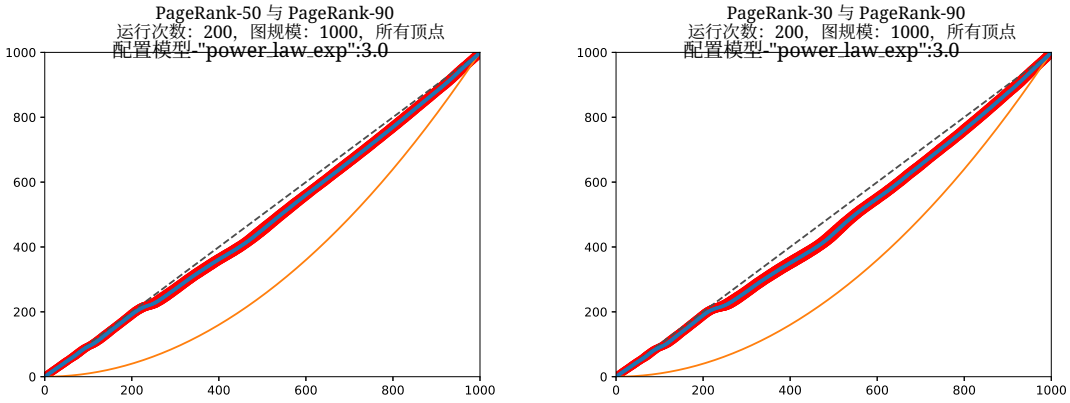

FIGURE 20. 人工网络中阻尼因子分别为 0.5 与 0.9、以及 0.3 与 0.9 时 PageRank 的 CCC。

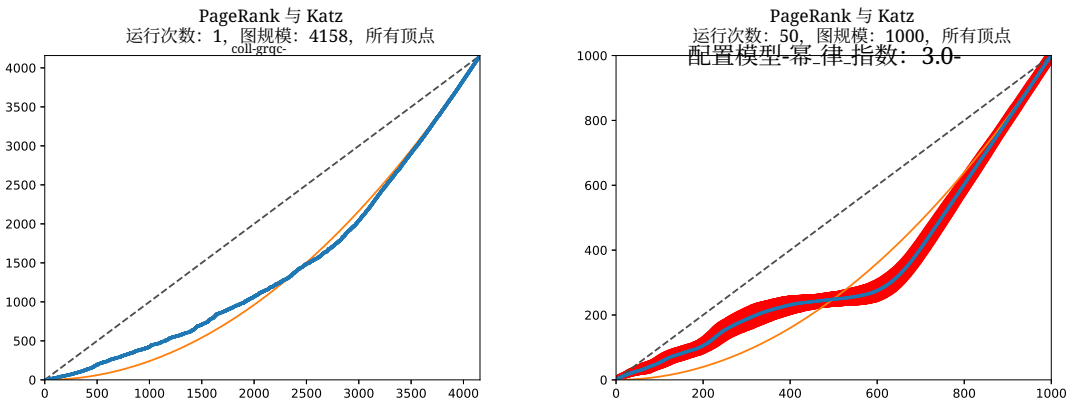

FIGURE 21. 合作网络 [51] 和人工网络中 PageRank 与 Katz 的 CCC。

A.4. **相关有向现实世界网络的 CCC**。本小节展示另外两个大型有向现实世界网络的 CCC 图。这两个网络在结构上不同于正文中讨论的引文网络，分别是斯坦福网络图和美国专利引文网络。其目的是说明，CCC 可以揭示中心性度量之间的强对齐关系，也可以揭示它们之间的近似独立性，具体取决于数据集和度量对。

斯坦福网络图。对于斯坦福数据集，介数与接近中心性的 CCC 靠近独立性曲线 x^2，这表明这两种度量识别出的排名靠前顶点集合在很大程度上互不重叠。介数与调和中心性、接近中心性与 PageRank，以及调和中心性与 PageRank 这几组度量也呈现出类似的近似独立性。

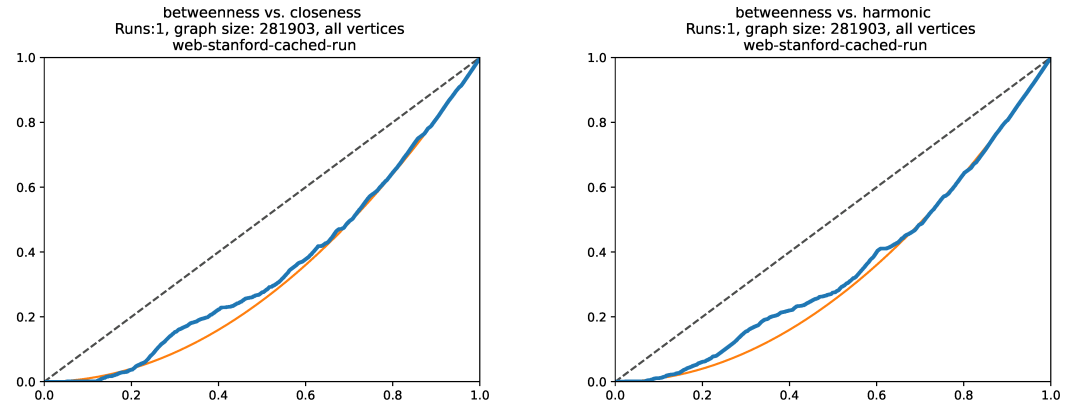

FIGURE 22. 斯坦福网络图：介数与接近中心性（左）以及介数与调和中心性（右）的 CCC。两条曲线都接近 x^2，表明它们近似独立。

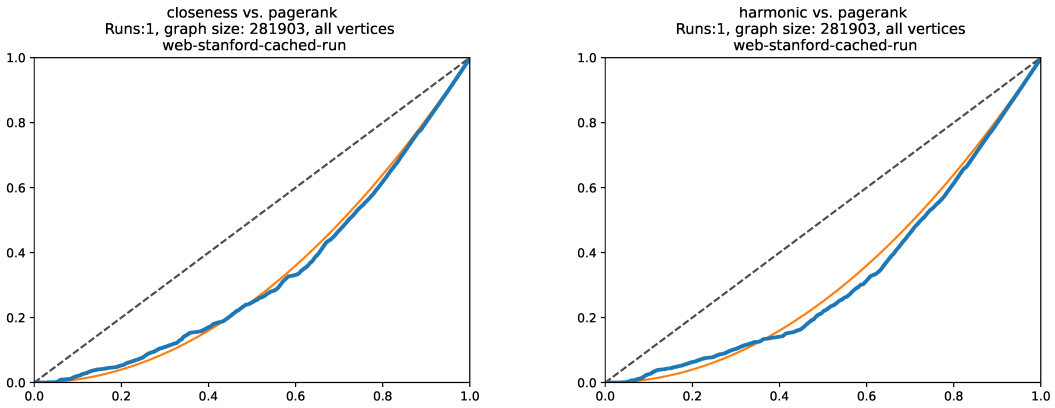

FIGURE 23. 斯坦福网络图：接近中心性的 CCC 与 PageRank（左），以及调和中心性与 PageRank（右）。两条曲线都接近 x^2，表明二者近乎独立。

美国专利引文网络。对于美国专利引文网络，若干组有向中心性度量呈现出近似独立性。具体而言，入度与出度、出度与 PageRank，以及 Katz 与出度的 CCC 均接近 x^2。这表明这些度量组合识别出的排名靠前顶点集合大不相同，凸显了在该数据集中，入链结构与出链结构所体现的重要性概念截然不同。

相比之下，PageRank 和入度表现出显著不同的行为。它们的 CCC 显著高于 x^2，表明存在很强的依赖关系。

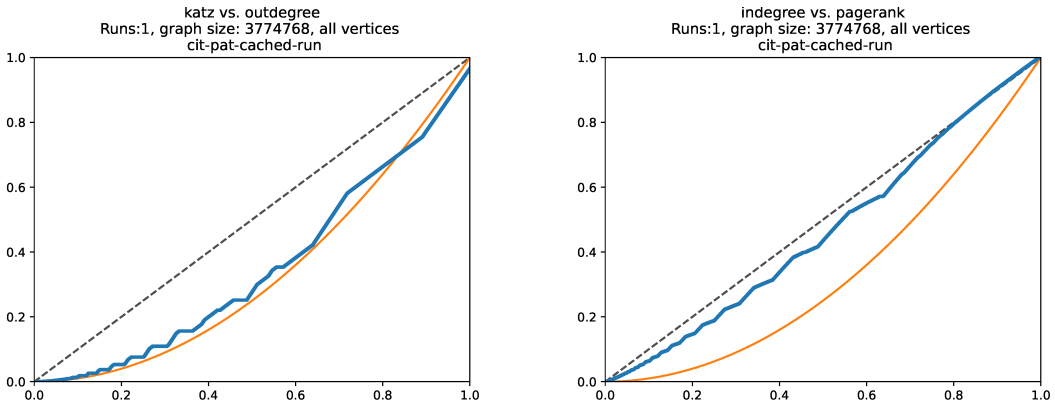

FIGURE 24. 专利网络：Katz 与出度的 CCC（左），以及 PageRank 与入度的 CCC（右）。Katz 与出度几乎相互独立，而 PageRank 与入度则表现出很强的依赖性。

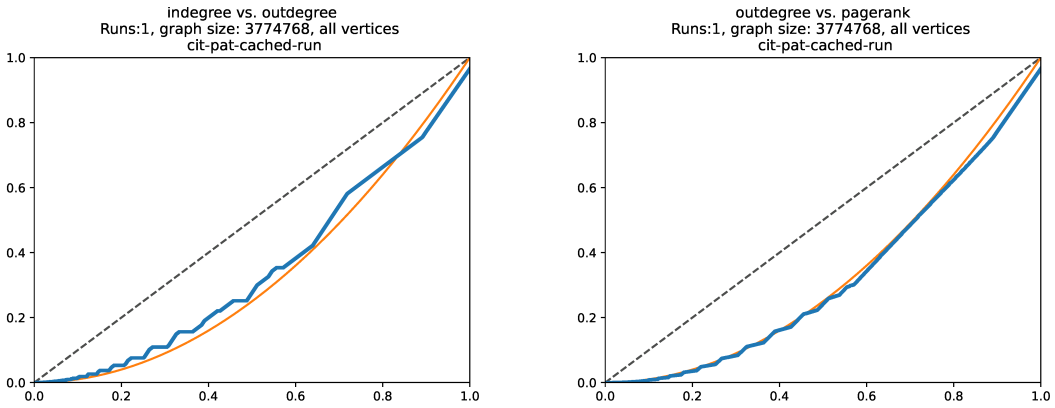

FIGURE 25. 专利网络：入度与出度的 CCC（左）以及出度与 PageRank 的 CCC（右）。两条曲线均接近 x^2，表明几乎具有独立性。

Email address: g.exarchakos@tue.nl

DEPARTMENT OF ELECTRICAL ENGINEERING, EINDHOVEN UNIVERSITY OF TECHNOLOGY

Email address: r.w.v.d.hofstad@tue.nl

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY

Email address: nagy@sdf.org

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY

Email address: m.pandey@math.au.dk

DEPARTMENT OF MATHEMATICS, AARHUS UNIVERSITY